CFDP Performance over Weather-Dependent Ka-band Channel

Sung I. U* and Jay L. Gao†
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109

This study presents an analysis of the delay performance of the CCSDS File Delivery Protocol (CFDP) over weather-dependent Ka-band channel. The Ka-band channel condition is determined by the strength of the atmospheric noise temperature, which is weather dependent. Noise temperature data collected from the Deep Space Network (DSN) Madrid site is used to characterize the correlations between good and bad channel states in a two-state Markov model. Specifically, the probability distribution of file delivery latency using the CFDP deferred Negative Acknowledgement (NAK) mode is derived and quantified. Deep space communication scenarios with different file sizes and bit error rates (BERs) are studied and compared. Furthermore, we also examine the sensitivity of our analysis with respect to different data sampling methods. Our analysis shows that while the weather-dependent channel only results in fairly small increases in the average number of CFDP retransmissions required, the maximum number of transmissions required to complete 99 percentile, on the other hand, is significantly larger for the weather-dependent channel due to the significant correlation of poor weather states.

Nomenclature

ARQ	=	Automatic Repeat ReQuest
AWGN	=	Additive White Gaussian Noise
B	=	bad weather state
BER	=	bit Error Rate
CFDP	=	CCSDS File Delivery Protocol
D_f	=	File delivery latency
DSN	=	Deep Space Network
EOF	=	end of file
F	=	exact probability distribution function
F_{wn}	=	CFDP delay performance on the *n*th weather sequences
FIN	=	finished message
FTP	=	File Transfer Protocol
G	=	good weather state
MRO	=	Mars Reconnaissance Orbiter
m	=	index in the cumulative distribution function
N	=	number of Protocol Data Units in the file
n	=	number of simulations need to run
NACK	=	negative Acknowledgement
N_s	=	number of the spurts required to transmit the file successfully
$N_{tx}(i)$	=	number of transmissions required to send the *i*th Protocol Data Unit
p	=	frame error rate in the good state

*Member of the technical staff, Communication Architecture and Research Section of the Jet Propulsion Laboratory, 4800 Oak Grove Drive M/S 238-420, Pasadena, CA 91109.
† Member of the research staff, Communication Architecture and Research Section of the Jet Propulsion Laboratory, 4800 Oak Grove Drive M/S 238-420, Pasadena, CA 91109.

q = frame error rate in the bad state
S_n = empirical average of the n independent random variables
T_{th} = noise temperature threshold
W = weather state
w_k = kth weather pattern
α = bit error rate in the good state
β = bit error rate in the good state

I. Introduction

THE success of Mars Reconnaissance Orbiter (MRO) insertion on March 10, 2006 marked a key milestone in the continuation of the Mars Networks, which now consists of several orbiters, each performing science as well as relay services for surface assets on the red planet. In particular, MRO will demonstrate high capacity communications over Ka-Band (32 GHz) which will greatly increase the capacity of the Mars Networks. The higher capacity is critical for delivering short turn-around, operational data as well as delay tolerant bulk science telemetry. However, data transmission in Ka-Band is highly vulnerable to weather impairments. Water vapor in the atmosphere will attenuate and radiate noise that causes packet errors. In addition to the weather effects, long propagation delay in the space communication links causes further loss in signal strength and reduces the remote transmitter's the ability to adapt to dynamics in the weather system.

In order to transfer files efficiently and reliably over long propagation space links, the CCSDS File Delivery Protocol (CFDP) [1] is designed with Automatic Repeat ReQuest (ARQ) retransmission mechanism. Unlike a terrestrial File Transfer Protocol (FTP), the CFDP has four error control modes to handle the link disruptions and outages frequently encountered in space [1].

The usage of the CFDP retransmission mechanism is expected to take advantage of the high capacity of the Ka-band channel to provide the low latency file transfer with automated reliability. Although analysis of the CFDP has been studied in [2],[3],and [4], the majority of these studies have not jointly considered the protocol interaction with the presence of correlated channel degradation and outages. For example, the CFDP delay was analyzed with the independent channel errors, however, weather effect was not considered in [2]. [3] has extensive discussions of the ka-band weather phenomena, but the protocol was not incorporated. In [4], a burst error channel model was analyzed; however, it did not address the delay performance of any higher layer protocols. [12] analyzed the bursty correlated channel only in the terrestrial communication environment. Therefore, there is a strong need to accurately predict the CFDP file latency performance under the weather-dependent Ka-Band channel and this will be analyzed in the present study with actual DSN data.

This paper is organized in the following way. First, the overall goal and approach of our study is briefly explained. Then, the CFDP retransmission mechanism and our modeling assumptions are presented. We then present different data sampling methods to gather weather statistics. Finally, we present analytical expression for the cumulative distribution function (CDF) of the number of transmissions required to complete file transfer and quantitative discussion on the average and the maximum latency under various scenarios.

II. Approach and Assumptions

Although accurate characterization of the weather effects is not a trivial task, a simple modeling approach can nonetheless be taken to capture the essential aspect of the ka-band channel and the CFDP performance by assuming that the atmospheric noise temperature is the primary source for communication errors [5]. Under this simple assumption, we ignored other factors that may contribute to the degradation of system performance such as antenna pointing loss, which in practice must be mitigated by using extra link margin. By modeling the link condition as solely driven by the atmospheric noise temperature, we isolate our analysis on the dependency between the CFDP and the weather correlation of the physical link.

In our model, we define a noise temperature threshold, T_{th}, that distinguishes the "good" and "bad" weather conditions. If the noise temperature is higher than the threshold, in a bad weather condition, significant error will occur. When the noise temperature is less than the threshold, in a good weather state, most of the transmitted

packets will be received successfully. Naturally, in the good weather period, only small random bit error rate (BER) is assumed. On the other hand, relatively high error rate are applied to capture severe packet losses during the bad weather condition.

To analytically derive the upper bound of the delay performance of the CFDP, a two-state Markov chain is exploited to capture the weather correlations. Actual temperature data collected from the DSN Madrid site are used to find the weather statistics. The extensive use of such model in existing research literature makes it an attractive first-cut approach for which allows mathematical analysis feasible [11].

There are several limitations and assumptions made throughout this work. We found that the duration of good and bad weather conditions do not fit exactly with a memoryless geometric random variable distribution, which will be shown in the later section. Therefore, actual weather data is not truly represented by the simple first-order Markov model. In general, a multi-state Markov model that captures finer granularity both in terms of channel state and long term correlation will be required for the precise modeling. Monte Carlo simulation of the protocol logic, rather than mathematical analysis, is necessary to quantify the multi-state Markov chain. However, we use two-state Markov chain for a first-cut modeling approach to get a high level understanding of delay performance. Also, a simple two-state Markov chain allows us to derive the CDF, and provides the theoretical foundation to analyze both the weather correlation and the protocol.

The focus of this work is to analyze the CFDP file latency at the link layer level rather than the bit level. Thus, we will not consider analyzing the received signal strength at the physical layer. Further, to model the performance at the link layer level, we resample the weather data on the scale of a round-trip time from Mars to DSN sites rather than the full resolution of the raw data, which is about 1.44 minute time scale. Since the CFDP retransmission process, as illustrated later, occurs on round-trip time scale, we condensed the noise temperature measurements to every round-trip time. That means that some dynamics in the original temperature data set will be lost. However, the following three sampling methods are explored to evaluate the impact of using the data on coarser time-scale: 1) sampling based on the average value, 2) the maximum value, and 3) the minimum value within each round-trip time interval.

III. Channel Model

In this study, we use a variation of the additive white Gaussian noise (AWGN) channel [6] to model different bit error rate (BER) in each weather state. This type of channel is also called the Gilbert-Elliot channel [7]. The channel has two weather states, a good and bad weather state, separated by the threshold value. During good weather conditions, most of the transmitted packets will be received successfully. During bad weather conditions, however, most of the transmitted packets will experience errors due to the high noise temperature at the receiving antenna. Therefore, two different BERs are applying to each good and bad weather state, as shown in Fig. 1. (b). We define the relatively high BER value for the bad weather state and fairly low BER value for the good weather state.

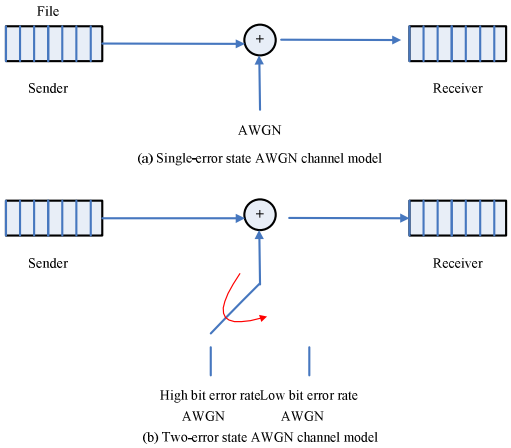


Figure 1. (a) Single error state AWGN channel model (b) two-error state AWGN channel model.

In the two-error state channel model in Fig. 1. (b), we assume that in a given weather state noise affects each transmitted packet identically and independently. This assumption is true when the file transmission time is short, compare to the changes in weather conditions. If we consider files that are at most 10MB in size and data rate about

1Mbps for ka-band, then the transmission time is about 80 seconds. As it has been shown in [5], the sample time constant is on the scale of 40 minutes. We believe it is a fair modeling approach to assume constant BER in each weather state. However, it should be noted that effects of antenna pointing error and the spacecraft elevation and data rate profile during a pass will introduce variations in signal strength even under constant weather, so that a fix BER model is only an abstraction of the practical condition. Nonetheless, we believe the independent packet error model is sufficient for analysis, providing a high level understanding of the CFDP performance over the ka-band channel.

The interactions between weather correlations and the CFDP are illustrated in Fig. 2, which depicts MRO sending files using the CFDP over ka-band. Between each transmission/retransmission attempt, the weather state undergoes stochastic transitions. One realization of the weather pattern is a {good, bad, bad, good} shown in Fig. 2. In each given weather state, noise affects each transmitted packets. The probability of success and the number of transmission attempts required to complete a file transfer strongly depends on the BER and the persistence of good and bad weather states. As our analysis will show, the "burstiness" of the good and bad weather states has significant impact on the spread, or deviation, of the probability distribution of the CFDP latency.

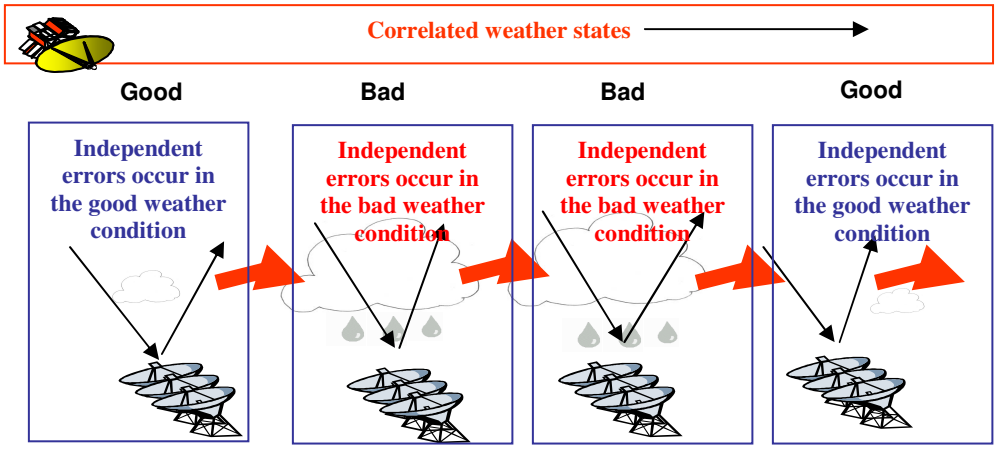
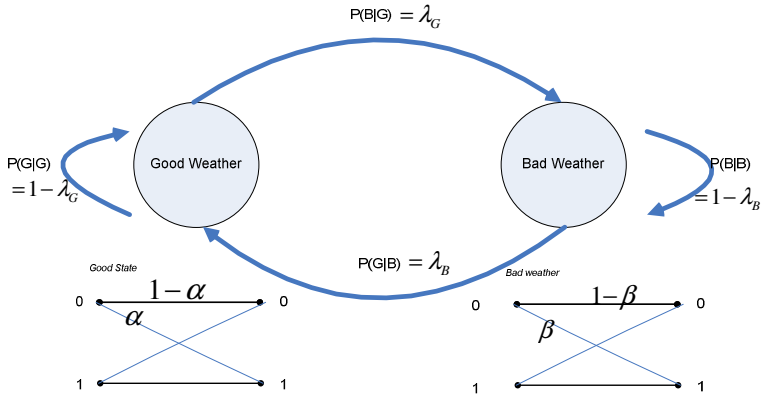


Figure 2. Example of the file transmission from the MRO to the DSN under different weather conditions.

To capture the weather correlation, the Gilbert-Elliot channel with two weather states are shown in Fig. 3. (a), which is essentially same as Fig. 1. (b). The transition from one state to another state is defined by the transition matrix, *P*, in Fig. 3.(b), which completely characterizes the channel behavior. In this model, the current state is determined by the previous state and λ_G and λ_B are the transition probabilities from good to bad and from bad to good state, respectively.



(a) Gilbert-Elliot channel

$$P = \begin{bmatrix} P(G|G) & P(B|G) \\ P(G|B) & P(B|B) \end{bmatrix} = \begin{bmatrix} 1-\lambda_G & \lambda_G \\ \lambda_B & 1-\lambda_B \end{bmatrix}.$$

(b) Transition matrix

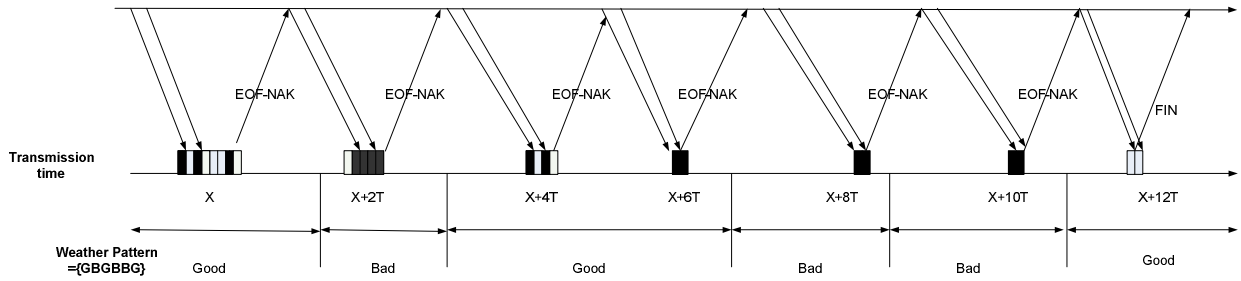
Figure 3. (a) Gilbert-Elliot channel model with a cross over probability α and β for the good and bad weather state (b) Transition matrix *P*.

The Gilbert-Elliot channel has a characteristic that the duration of being in a good and bad weather state is geometrically distributed with mean, $1/\lambda_G$ and $1/\lambda_B$ [9]. Also, there is an independent crossover probability α and β that corresponds to the BER associated with each weather state.

IV. CCSDS File Delivery Protocol (CFDP)

For the CFDP file transfer, each file is divided into multiple protocol data units (PDU)s marked with a sequence number. The end of file PDU (EOF-PDU) is sent at the end of the initial file transmission attempt. The receiving CFDP entity, after receiving the EOF-PDU, will reply, if necessary, with a NAK message listing all the PDUs that were not successfully received during the initial transmission attempt. Upon the reception of this NAK message, the sender side retransmits the PDUs and the receiver will again responds with NAK messages if further retransmission is needed. This process will repeat with periodicity approximately equal to the round-trip time until all PDUs are received correctly; if all PDUs are received correctly, the receiver will send finished message (FIN) to the sender and close the file transmission. Each round of interaction between the sender and the receiver is sometimes referred to as a 'spurt'. The number of spurts required to complete a file transfer is a random variable, depending on the error characteristics of the channel and the file size [2]. The number of spurts basically determines the number of round-trip messaging required and, therefore, it captures the file transfer latency.

In this study, we assume the NAK message on the uplink is very reliable, and ignore the effects of protocol timers as defined in the original deferred NAK mode [1]. The timers are typically associated with the loss of the EOF and NAK messages, which we ignore in order to simplify our analysis. In practice, it is fair to assume that NAK transmission are very robust if X-band uplink is used. However, the loss of EOF is, on the other hand, subject to the same weather pattern as the data PDUs and therefore should be taken into account in follow-on studies. To better understand the CFDP operation, following example is given:



- Weather pattern {GBGBBG}
- *T* : One-way trip time (could be as long as 20 minutes)
- *X* : Time when first PDU is received at the DSN station (min)
- File has 8 PDUs
- File transfer took 7 spurts (transmissions) to complete
- Each black marked PDU is the one needs to be retransmitted in the next pass

Figure 4. Example of a CFDP file transmission

In Fig. 4, with the given weather pattern {GBGBBG}, *T* is defined as an one-way trip time from the sender to the receiver. *X* represents the time of initial PDU receptions. In this example, the file is composed of 8 PDUs and 7 spurts are required to receive all PDUs correctly.

V. Weather statistic and sampling

The atmospheric noise temperatures are collected at the DSN Madrid for 5,356 days. There are some inconsistencies and gaps in the data set; hence, preprocessing is required to acquire a subset of data with a consistent time interval between each data point. After preprocessing, 1,956 days of data consist of 1,532,456 temperature measurements are obtained where every data point is almost 1.44 minutes apart.

Sampling

Due to the long propagation delay, the time difference between the previous file transmission attempt and the current file transmission attempt is relatively large compared to the original noise temperature data collection time interval. For the link layer perspective, it is more meaningful to get the weather statistics based on every round-trip time interval than 1.44 min interval. The weather condition at the instance of PDUs' arrival at the DSN antenna is the most important time for PDUs' successful receptions. Therefore, there is a strong need to resample data in every round-trip time than simply using the 1.44 min interval. Specifically, we assume a 40 min round-trip time from Mars to the DSN, so the factor by which the data set is condensed is given by:

$$\text{Number of data points} = \frac{\text{Round-trip time}}{\text{Sample interval}} = \frac{40 \text{ min}}{1.44 \text{ min}} = 27.78. \quad (1)$$

As shown in Eq. (1), we need to sample once every 28 data points in the original data set. However, by sampling, some of weather information will be lost. In order to capture the envelop on the impact of under-sampling the data, we processed the data in three different ways: 1) average every 28 samples and use the average value, 2) select the maximum reading among the 28 samples, and 3) choose the minimum reading of the 28 samples. The reasoning behind averaging 28 samples is to compress and compact the original data points, while still capturing the relative contribution of each data points; choosing the maximum and minimum sample should provide the upper and the lower bound to evaluate the potential deviation due to under-sampling of the data set.

Once temperature data is sampled to round-trip time scale, a threshold, T_{th}, can be defined in the following way to convert the temperature data into two groups:

$$\begin{cases} T > T_{th}, \text{Bad weather}, (W = B) \\ T \le T_{th}, \text{Good weather}, (W = G). \end{cases} \quad (2)$$

If the sampled noise temperature is greater than the threshold, we define the weather as being in the bad state, and denote it as, $W = B$. Otherwise, we define weather is in the good state and represent it as, $W = G$. The statistics of the sampled data by averaging, maxima, and minima are shown in Table 1. with the threshold of 20 *K*.

To describe the probability of encountering good weather condition, *availability* metric is defined as the percentage of time it is capable of providing services [8]. Thus, availability can be calculated as a percentage of good weather in the preprocessed DSN temperature data and the 20*K* threshold roughly corresponds to the 89% weather availability shown in Table 1. With a 20*K* threshold, the statistics of the temperature data obtained from three different sampling methods are recorded in Table 1.

From Table 1, we found that the sampling with averaging produces a fair match with the availability and mean obtained from the original weather data set. Sampling the maximum and minimum data points causd significant deviation from the statistics of the original data set shown in Table 1.

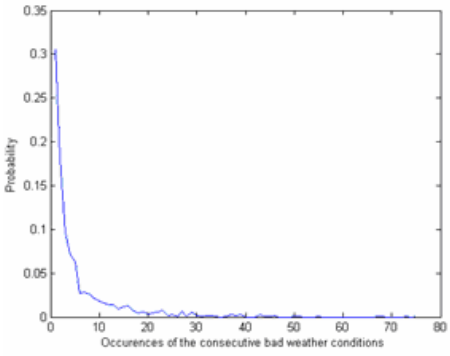
Table 1. Statistics of the temperature data sampled in every round-trip time interval with the threshold = 20k.

	Preprocessed original data	**Sampling data by averaging**	**Sampling data by selecting max**	**Sampling data by selecting min**
Availability	89.42%	88.04%	82.47%	92.54%
Data size(Number of samples)	1532456	54730	54730	54730
Max(K)	270.5059	214.76	270.5059	159.809
Min(K)	5.3324	5.9289	6.4332	5.3224
Mean(K)	16.1889	16.19	19.6919	13.9116
Std	14.8673	14.1322	22.2187	9.5796

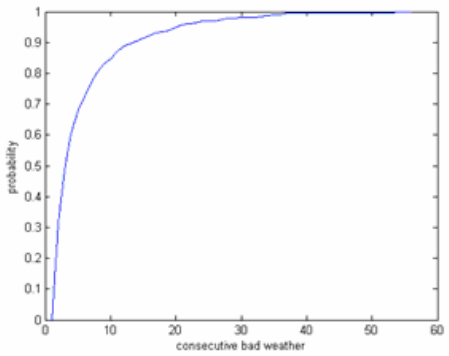
VI. Estimating Parameters from Sample Statistics

Good and bad weather processes

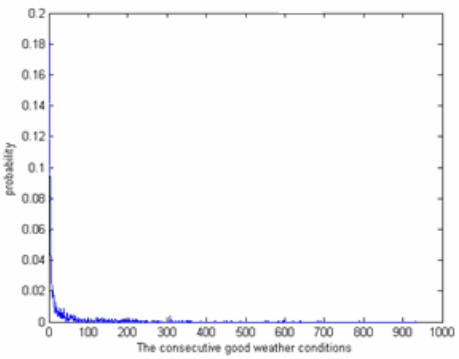
Once weather data is sampled in every round-trip time, the average duration of the consecutive good and bad weather are tabulated. Our goal is to approximate the lengths of the consecutive good and bad weather as a memoryless geometric random variable so that a two-state Markov chain can be applied. With data obtained from three different sampling methods, the probability density function (PDF) and the cumulative distribution function (CDF) of the number of consecutive good and bad weathers are plotted. The PDF and CDF from the average data sampling are shown in Fig. 5. In Fig. 6.(c), the theoretical CDF of the geometric random variable with the mean $(1/\lambda)$ obtained in Table 1 is overlaid with the CDF obtained from the sampled weather statistic. It is noted that these two CDFs do not align very well; there are noticeable deviations, especially for the good weather distribution. Therefore, to accurately model the performance of the CFDP, more states will typically be required. However, the geometric random variable assumption provides the simple way to analyze the complicate weather events. Therefore, in the present scope of the study, we model the good and bad weather processes with a geometric distribution.



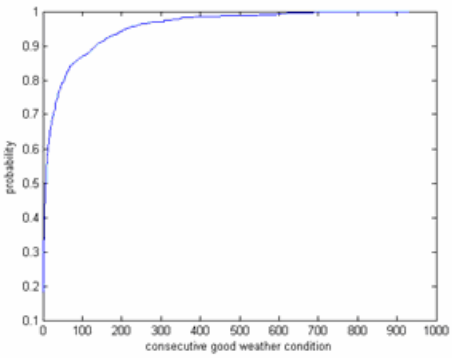
(a) Bad weather PDF



(b) Bad weather CDF

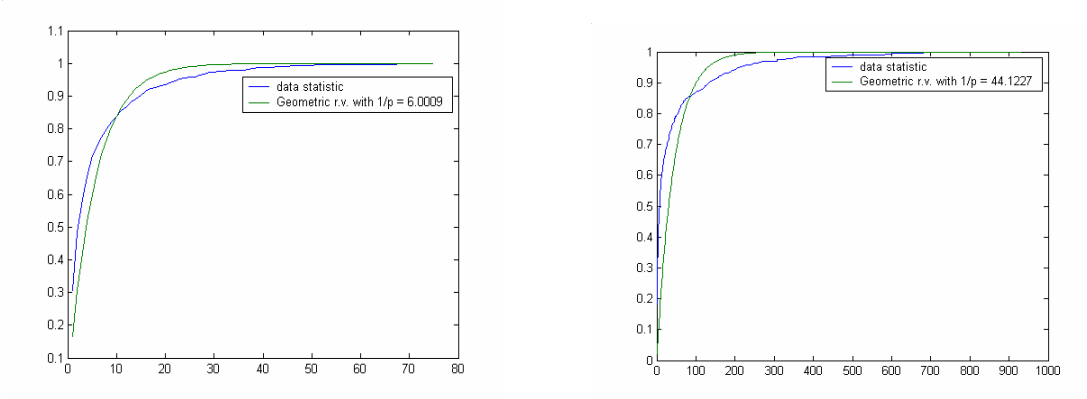


(c) Good weather PDF



(d) Good weather CDF

Figure 5. The PDF and CDF plot with threshold = 20K.



(a) Bad weather **(b) Good weather**

Figure 6. Comparisons between the CDF obtained from actual data samples and the geometric random variable with mean obtained from weather statistic in Table 1.

The duration of good and bad weather statistic by averaging, maxima, and minima with a threshold of 20*K* are recorded in Table 2. The maximum, minimum, and mean duration of the consecutive good and bad weather periods are calculated. As expected, sampling by selecting the maximum yields the longest average duration of the bad weather and the shortest average good weather duration. Opposites are true with the minimum data sampling method, as expected. The statistics obtained from sampling by averaging reside between the two extremes. The data obtained from Table 2. is consistent with the result from Table 1. From Table 1, we can observe that there are more statistical variations in the good weather distributions. This is partly because most of temperature data are below 20*K*, which are translated into good weather data points. Hence, how we sample the original data can alter the original good weather data statistic. On the other hand, there are quite less number of bad weather data points; therefore, bad weather statistic is less sensitive to the choice of data sampling method. The relationship between the selection of data sampling methods and the delay performance are evaluated in the later section.

Table 2. Duration of bad and good weather statistic by averaging, choosing maximum and minimum sampling with threshold = 20 *K*.

	(Bad weather)			(Good weather)		
	Sampling by averaging	**Sampling by choosing the max**	**Sampling by choosing the min**	**Sampling by averaging**	**Sampling by choosing the max**	**Sampling by choosing the min**
Max duration	75	111	74	932	600	1771
Min duration	1	1	1	1	1	1
Mean duration	6.0009	6.1817	5.5586	44.1227	29.06	68.1902
Std	8.7483	9.9479	8.0311	95.5634	66.1582	162.1346

As defined in the earlier section, the transition probability from good to bad states can be derived by 1 / (average duration of good weather) and the transition probability from bad to good state is 1 / (average duration of bad weather). The transition probabilities of different sampling methods with the threshold of 20*K*, are shown in the following Table 3. These transition probabilities will be used for simulating weather patterns.

Table 3. Transition probabilities with threshold = 20*K*.

	Sampling by averaging	**Sampling by choosing the max**	**Sampling by choosing the min**
P(G|G)	0.9773	0.9656	0.9853
P(B|G)	0.0227	0.0344	0.0147
P(G|B)	0.1667	0.1618	0.1799
P(B|B)	0.8333	0.8382	0.8201

VII. Mathematical Analysis

In this section, the expression of the CDF of the number spurts required to complete the file transmissions is derived and evaluated. The CFDP file delivery latency, D_f, is given by the

$$D_f = T \cdot (1 + 2(N_s - 1)), \tag{3}$$

where *T* is the one-way propagation delay and N_s is the random variable representing the number of spurts required to transfer a file completely and correctly.

Let *N* be the number of PDUs in the file, $N_{tx}(i)$ be the number of transmissions required to send the *i*th PDU, and w_k be the *k*th weather pattern, which is a single realization of weather pattern to measure the CFDP file transfer until success. Then, the CDF of N_s is defined as follows:

$$P[N_s \le m] = \sum_{k=1} P[N_s \le m \mid w_k] \cdot P[w_k], \tag{4}$$

where $w_1, w_2, \ldots w_k, \ldots$ are mutually exclusive weather patterns. By applying the theorem of total probability [9], Equation (4) can be obtained.

However, evaluating the above expression over all possible realizations of weather patterns cannot be done manually for the potentially large weather patterns. Therefore, we rely on numerical methods to generate the realizations of weather patterns, where each weather pattern is produced according to the transition probabilities in Table 2.

Once weather pattern is generated, the analysis follows similarly with the independent channel in [2]. For a given weather pattern, the noise affects each PDU independently. Therefore, the number of transmissions required to deliver the *i*th PDU, $N_{tx}(i)$, is independent with the transmission of the *j*th PDU, $N_{tx}(j)$.

It is observed that the number of spurts required to complete a file transfer, N_s, is simply the maximum number of transmissions required among all PDUs in the file. Therefore, N_s can be expressed in the following way:

$$N_s = \max_{i \in \{1,\ldots,N\}} \{N_{tx}(i)\}. \tag{5}$$

By substituting Eq. (5) into Eq. (4), following equation can be obtained,

$$P[N_s \le m \mid w_k] = P[\max_{i \in \{1,\ldots N\}} \{N_{tx}(i)\} \le m \mid w_k]. \tag{6}$$

The noise affects each PDU independently in the given weather patterns, Eq. (6) becomes

$$\begin{aligned} P[\max_{i \in \{1,\ldots N\}} \{N_{tx}(i)\} \le m \mid w_k] &= P[N_{tx}(1) \le m \mid w_k] \cdot P[N_{tx}(2) \le m \mid w_k] \cdots \\ P[N_{tx}(i) \le m \mid w_k] &\cdots P[N_{tx}(N) \le m \mid w_k]. \end{aligned} \tag{7}$$

Since $N_{tx}(i)$ is independently identically distributed (i.i.d.), Equation (7) can be written into the product form as follows:

$$\begin{aligned} &\prod_{i=1}^{N} P[N_{tx}(i) \le m \mid w_k] \\ &= \left\{ P[N_{tx}(1) \le m \mid w_k] \right\}^N \\ &= \left\{ \sum_{j=1}^{m} P[N_{tx}(1) = j \mid w_k] \right\}^N. \end{aligned} \tag{8}$$

If we expand each term in Eq. (8), we obtain

$$\left\{ P[N_{tx}(1) = 1 \mid w_k] + \ldots + P[N_{tx}(1) = i \mid w_k] + P[N_{tx}(1) = i+1 \mid w_k] + \ldots + P[N_{tx}(1) = m \mid w_k] \right\}^N. \tag{9}$$

Let us introduce

$$P(i,k) = P[N_{tx}(1) = i \mid w_k], \tag{10}$$

as the probability of *1*st PDU transmission being successful in the *i*th transmission attempt given the *k*th weather pattern. Using Eq. (10), Equation (8) can be rewritten as

$$\left\{\sum_{j=1}^{m} P[N_{tx}(1) = j \mid w_k]\right\}^N = \left\{P(1,k) + P(2,k)... + P(i,k) + P(i+1,k) + ... + P(m,k)\right\}^N. \tag{11}$$

$P(i+1,k)$ in Eq. (11) can be computed from the following recursive relation,

$$P(i,k) = \begin{cases} P'(i,k)\cdot(1-p), w_k(i) = G \\ P'(i,k)\cdot(1-q), w_k(i) = B \end{cases},$$

$$P'(i,k) = \begin{cases} P(i,k)\cdot\dfrac{p}{1-p}, w_k(i) = G \\ P(i,k)\cdot\dfrac{q}{1-q}, w_k(i) = B \end{cases}, \tag{12}$$

$$P(i+1,k) = \begin{cases} P'(i,k)\cdot(1-p), w_k(i+1) = G \\ P'(i,k)\cdot(1-q), w_k(i+1) = B \end{cases},$$

where *i*, and *k* are the positive integer and $w_k(i)$ is the *i*th weather condition in the *k*th weather pattern, and p and q are the PDU error rates for good and bad weather. $P'(i,k)$ is the probability of transmission failure up to *i*th transmission attempts in the *k*th weather sequence. $P'(i,k)$ is, therefore, the intermediate term to calculate the next $P(i+1,k)$ term through the recursive relationship, assuming the *i*th PDU transmission has been failed .

Using Eq. (12), each term in Eq. (11) can be calculated in the following way,

$$P(1,k) = \begin{cases} (1-p), w_k(1) = G \\ (1-q), w_k(1) = B \end{cases},$$

$$P'(1,k) = \begin{cases} P(1,k)/\dfrac{p}{1-p}, w_k(1) = G \\ P(1,k)/\dfrac{q}{1-q}, w_k(1) = B \end{cases},$$

$$P(2,k) = \begin{cases} P'(1,k)\cdot(1-p), w_k(2) = G \\ P'(1,k)\cdot(1-q), w_k(2) = B \end{cases}, \tag{13}$$

…

Let us introduce $f(i)$, and $g(i)$ to further simplify the terms in Eq. (13),

$$f(i) = \begin{cases} \dfrac{p}{1-p}, w_k(i) = G \\ \dfrac{q}{1-q}, w_k(i) = B \end{cases}, \; g(i) = \begin{cases} 1-p, w_k(i) = G \\ 1-q, w_k(i) = B \end{cases}. \tag{14}$$

Note that $f(i)$ and $g(i)$ will greatly simplify Eq. (11) with single weather state case, which is proved in Appendix section.

Using Eq. (14), Equation (11) can be written in the functional form as follows:

$$P[N_s \le m \mid w_k] = \{P(1,k) + P(2,k) + ... + P(i,k) + P(i+1,k) + ... + P(m,k)\}^N$$
$$= \{P(1,k) + P'(1,k) \cdot g(2) + ... + P'(i-1,k) \cdot g(i) + P'(i,k) \cdot g(i+1) + ... + P'(m-1,k) \cdot g(m)\}^N \quad (15)$$
$$= \{P(1,1) + ... + P(1,1) \cdot f(1) \cdot g(2) \cdot ... \cdot f(i-1) \cdot g(i) + ... + P(1,1) \cdot f(1) \cdot g(2) \cdot ... \cdot f(m-1) \cdot g(m)\}^N .$$

Equation (15) is the expression of the CDF of the number of transmissions required to complete a file transfer over the *k*th weather sequences. To quantify the CDF, we apply a numerical computation method that iterates over a very large set of weather patterns.

The number of iterations required to obtain the reliable CDF can be found in Hoeffding's Inequality [10]. Hoeffding's Inequality in Eq. (16) provides the bound on the number of iterations required to achieve the true mean $E[F]$

$$P[|S_n - E[F]| \ge v] \le 2e^{(-2nv^2)}, \quad (16)$$

where F is the exact probability distribution function bounded in [0,1], n is the number of iterations needed to simulate, v is the threshold, and, S_n, is the ensemble average. In our case, let F_{wk} be the CFDP delay on the *k*th weather sequences. Then, the CDF of 1 through n realizations are expressed in the following way:

$$F_{w1} = \{P(1,1) + P(2,1)... + P(i,1) + P(i+1,1) + ... + P(m,1)\}^N$$
$$... \quad (17)$$
$$F_{wn} = \{P(1,n) + P(2,n)... + P(i,n) + P(i+1,n) + ... + P(m,n)\}^N .$$

Let us define the, S_n, as the ensemble average of all the n weather patterns,

$$S_n = \frac{F_{w1} + ... + F_{wn}}{n} . \quad (18)$$

By Hoeffding's Inequality, we can approximate the CDF if we find the proper n as follows,

$$P[N_s \le m] = \sum_{k=1}^{n} P[N_s \le m \mid w_k] P[w_k]$$
$$\approx S_n (\text{for large } n). \quad (19)$$

In this work, lengths of 1,000,000 random good and bad weather realizations are generated according to the transition probabilities found in Table 3. The CFDP simulation starts at a random location once the weather sequences are generated long enough to be in the stationary distribution. Four thousands different weather sequences are generated with the length of each sequence to be 1,000,000. The differences between the simulation results and the true mean are computed below for ε =0.1, 0.05, and 0.02,

$$P[|S_n - E[F]| \ge .1] \le 2e^{(-2n(.1)^2)} = 3.6 \cdot 10^{-35}$$
$$P[|S_n - E[F]| \ge .05] \le 2e^{(-2n(.05)^2)} = 4.12 \cdot 10^{-12} \quad (20)$$
$$P[|S_n - E[F]| \ge .02] \le 2e^{(-2n(.02)^2)} = 0.0815.$$

The probability of being differed by more than 0.02 between the empirical average and the true mean is 0.0815 for 4,000 simulation runs. Four thousands simulation runs seem to be good to achieve the accurate ensemble average of CDF to evaluate the CFDP delay performance.

VIII. Evaluation of the CFDP Latency

In this section, the CDF is evaluated over various BERs, and file sizes over three different data sets obtained from average, maxima, and minima sampling methods. Computation utilizing the parameters in Table 5 is conducted to evaluate the CDF of the CFDP file latency. We consider 40 min round-trip time from Mars to the DSN site, as described in the earlier section. A single PDU size is fixed as a 1KB with file size at either 1 or 10MB. BERs of 10^-5, 10^-6, 10^-7, and 10^-8 are considered for good weather state; 10^-3 and 10^-4 BERs are used for

the bad weather state. 10^-3 and 10^-4 BER corresponds to the 99 percent and the 55 percent of PDU error rate respectively, as shown in Table 6.

The overall goal of this evaluation is to find the CDF of the maximum number of spurts required to successfully complete the file transfer. The CDF plot captures the dynamics in the behavior of the CFDP over weather-dependent link. The CDF provides design guidelines for mission planners who wish to have quantitative understanding on percentile upper-bound on the file transfer latency when using the CFDP. In general, the higher the probability requirement, say 99.9% or 99.99%, the higher latency bound will be in order to guarantee the file completion. Besides the upper bound of the delay performance, the average number of transmissions required is another major interest because it represents the long term average performance of the system.

Table 5. The CDF evaluation parameters.

Simulation Parameters	Values
Round-trip time from Mars to Earth (min)	40 min
PDU size	1KB
BER in good weather	10^-5, 10^-6, 10^-7, 10^-8
BER in bad weather	10^-3, 10^-4
File size	1MB, 10MB
Availability	88 percent
Data sampling method	Average data sampling, Maximum data sampling, Minimum data sampling

Table 6. BER and the corresponding PDU error rate.

BER	PDU error rate
10^-3 (bad)	0.9997
10^-4 (bad)	0.5507
10^-5 (good)	0.0769
10^-6 (good)	0.008
10^-7 (good)	7.99968 x 10^-4
10^-8 (good)	7.9997 x 10^-5

A. Effect of applying different BERs in good and bad weather

In this section, the effects of using different BER on the fixed file size with data obtained from the average sampling methods are examined. Different combinations of good and bad weather BER pairs are evaluated and the results are shown in Fig. 7, and Fig. 8. To easily refer to the data points in the figures, we introduce a parenthesis notation to indicate the good and bad weather BER pair. For example, (*a*, *b*) indicates that *a* is the BER at good weather, and *b* is the BER at bad weather.

To achieve the 99% file completion, the maximum numbers of transmission required are 20, 17, 16, and 15 for (10^-5, 10^-3), (10^-6, 10^-3), (10^-7, 10^-3), and (10^-8, 10^-3) BER pairs, as shown in Fig. 7. Notice that reducing the BER in good weather condition enhances the required number of spurts by 25%, from 20 to 15 transmissions.

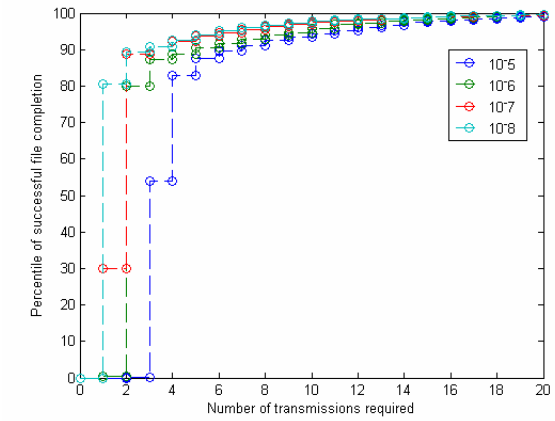


Figure 7. 10MB file latency performance with 10^-3 bad weather BER.

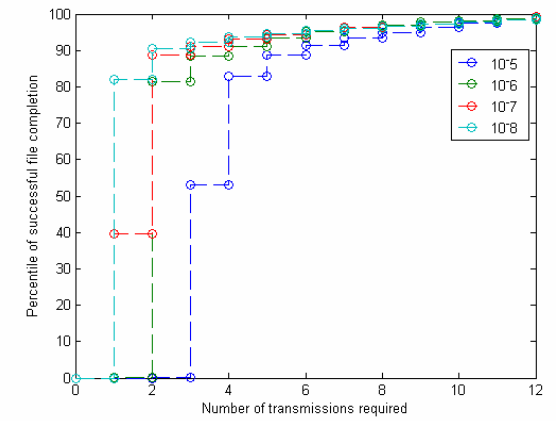


Figure 8. 10MB file latency performance with 10^-4 bad weather BER.

The result for 10^-4 bad weather BER are recorded in Fig. 8. The maximum number of transmissions required to complete the 99% of file transfers are all 12 for (10^-5, 10^-4), (10^-6, 10^-4), (10^-7, 10^-4), and (10^-8, 10^-4) BER pairs. From Fig. 7.with Fig. 8, the effect of BER in bad weather conditions can be examined. By reducing the

bad weather BER from 10^-3 to 10^-4, the maximum number of transmissions required to achieve the 99% file completion goes down from 20 to 12 transmissions with the 10^-5 good weather BER; the maximum number of transmission goes down from 15 to 12 transmissions with 10^-8 good weather BER. The gain is achieved from the fact that about half of PDUs are received correctly in 10^-4 bad weather BER case, whereas almost entire PDUs are lost in the 10^-3 bad weather BER.

Although the maximum number of transmissions required to achieve 99% file completion is more than 10, the average number of transmissions are much fewer. Using the following equation, the average number of transmissions required to complete the 99% of file transfer can be computed:

$$E[\text{number of transmissions}] = \sum_{k=1} P[k] \cdot k \,, \tag{21}$$

where the k is the number of transmissions required.

For 99% of the file completion, the average number of transmissions are 4.19, 2.91, 2.23, and 1.73 for (10^-5, 10^-3), (10^-6, 10^-3), (10^-7, 10^-3), and (10^-8, 10^-3) BER pairs, as shown in Fig. 9. By reducing the BER from 10^-5 to 10^-8 in good weather, at most 2.46 transmissions can be reduced on average, as shown in Fig. 9. For each (10^-5, 10^-4), (10^-6, 10^-4), (10^-7, 10^-4), and (10^-8, 10^-4) pair, the average number of transmissions are 3.92, 2.54, 2.00, and, 1.54.

By increasing the bad weather state BER from 10^-3 to 10^-4, about 0.27, 0.37, 0.23, and 0.19 transmissions are saved on average for each 10^-5, 10^-6, 10^-7, and, 10^-8 BER pair. As shown in Fig. 9, a moderate performance gain can be achieved on average by reducing a BER for both good and bad weather from (10^-5, 10^-3) to (10^-8, 10^-4), which saves up to 2 transmissions.

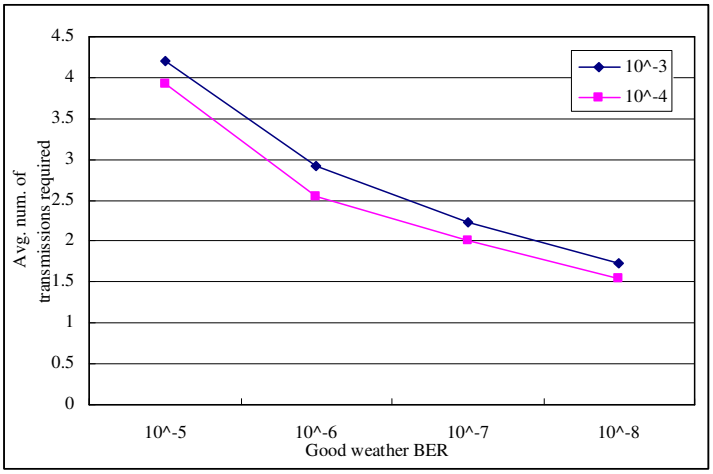


Figure 9. Average number of transmissions required to achieve the 99% file completion of 10MB file size.

We can observe that average numbers of transmissions required are a lot less than the maximum numbers of transmissions we computed. This is because bad weather patterns at the tail of the simulation are averaged out.

B. Effect of different file sizes

In this section, 1MB file size is considered with same parameters. The CDF of the number of transmissions required for 99% file completion are plotted in Fig. 10, and Fig. 11. From Fig. 10, and Fig. 11, we can see that the smaller file size improves latency performance. The order of improvement can be found by comparing Fig. 7, and Fig. 8. with Fig. 10, and Fig. 11. The 1MB file size saves up to two transmissions, or equivalently about two round-trip times, for 99% file completion.

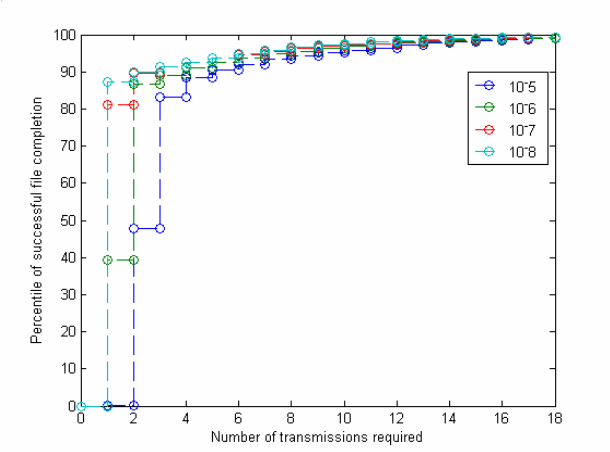


Figure 10. 1MB file latency performance with bad weather BER at 10^-3.

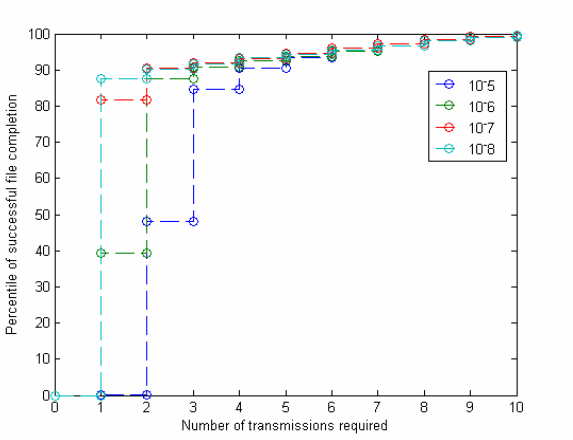


Figure 11. 1MB file latency performance with bad weather BER at 10^-4.

The average number of transmissions required is shown in Fig. 12. The differences of the average numbers of transmissions required with the 10MB and 1MB file sizes are plotted in Fig. 13. in the 10^-3 and 10^-4 bad weather BER. It is shown that, on average, about 3.2 transmissions and 1.5 transmissions are required to transmit 1MB file for 10^-3 and 10^-4 bad weather BER, respectively. From Fig. 13, we can see that reducing the file size has the performance gain ranging from 0.1 to 1 transmission. The reduction in average latency is quite significant, yielding more than 50% reduction.

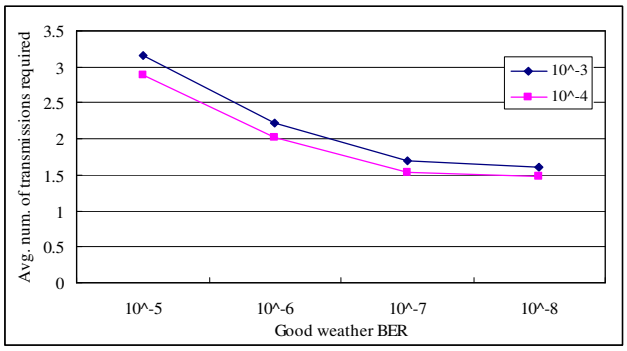


Figure 12. Average number of transmissions required to achieve the 99% file completion of the 1 MB file size.

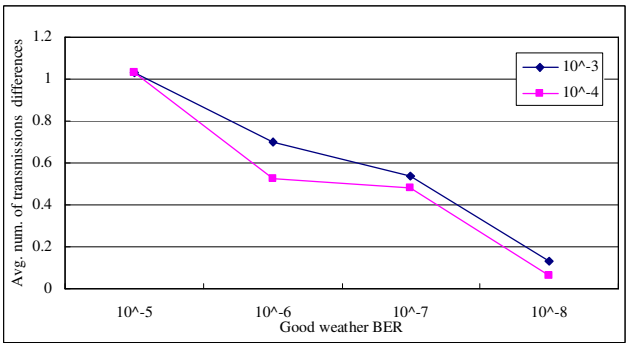


Figure 13. Differences between the average numbers of transmissions required to achieve the 99% of 10 MB and 1MB file with 10^-3 and 10^-4 bad weather BER.

Here it should be noted that the file size has more impact on the 99 percentile latency instead of the average latency performance. One possible explanation is that the 99 percentile latency computation has a stronger dependency on the weather correlation at the tail events; whereas the average latency does not reflect the correlation as strongly because the tail events are averaged out over the entire distribution. Therefore, this indicates that if the system is designed to meet the high percentile latency performance, one need to improve, on a fundamental level, the ka-band link's sensitivity to weather events using adaptive methods and weather forecasting as described in [5].

C. Effect on different sampling methods

In this section, data obtained from different sampling methods are compared in order to evaluate the sensitivity of our analysis. The range of statistical errors introduced by sampling a point from every 28 data samples is evaluated. We confirm that choosing the maximum among 28 samples provides the upper bound and choosing the minimum can provide the optimistic lower bound for actual latency.

First, computations were conducted using sampled data set obtained from the maximum sampling method. From Fig. 14, and Fig. 15, the maximum sampling method yields up to seven more transmissions for the 99% file completion than the transmissions required from the average sampling method shown in Fig, 7, and Fig. 8. This is because there are fewer good weather states in the data set obtained from the maximum sampling method. This is directly resulted from the low weather availability shown in Table 1. Specifically, seven more transmissions are required for (10^-5, 10^-3) BER pair, and four more transmissions for the (10^-5, 10^-4) BER pair compared to the result obtained from the average sampling method in Fig. 7, and Fig. 8.

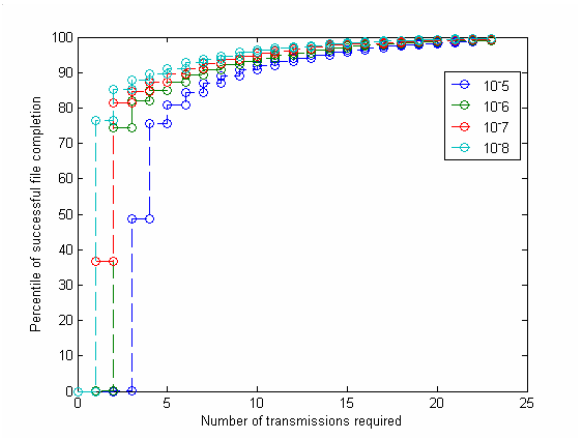


Figure 14. 10MB file latency performance with bad weather BER at 10^-3 with data obtained from maximum sampling.

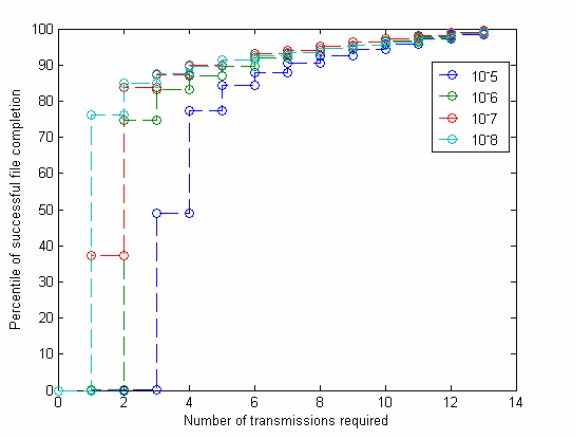


Figure 15. 10MB file latency performance with bad weather BER at 10^-4 with data obtained from maximum sampling.

The average number of transmissions required to complete the 99% of the 10 MB file size are recorded in Fig. 16. with 10^-3 and 10^-4 bad weather BERs. On average, about 4.7 transmissions are required to transmit the 10MB file for the (10^-5, 10^-3) BER pair and 2 transmissions for the (10^-8, 10^-4) BER pair.

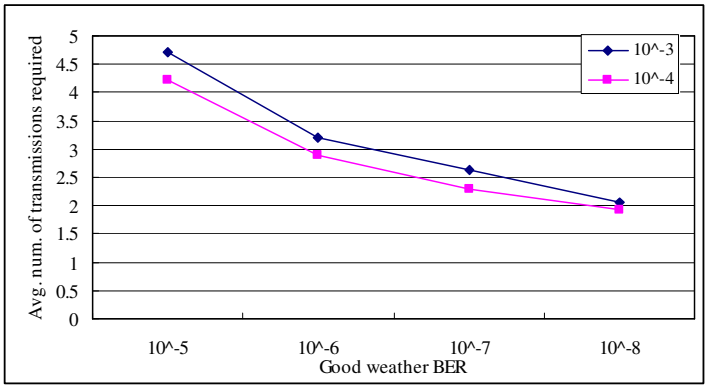


Figure 16. Average number of transmissions required to achieve 99% completion of the10 MB file on the data obtained from maximum sampling.

Also, the differences between the maximum sampling method and the average samplings are shown in Table 7. The deviations are around 0.5 or less transmissions between two different sampling methods in terms of average delay performance. Errors due to data alternations from the mis-sampling could be less than the 1 transmission difference. Hence, we can observe that mis-sampling the original data does not introduce the significant performance deviations.

Table 7. Differences between the average numbers of transmissions required to achieve the 99% completion of the 10 MB file from the maximum and the average sampling method.

BER	10^-5	10^-6	10^-7	10^-8
10^-3	0.5115	0.3003	0.4019	0.3301
10^-4	0.304	0.3376	0.2898	0.3848

Using the same parameters, the simulation is run with the data obtained from sampling minimum data point. The CDF of the number of transmissions required for the 99% file completion are shown in Fig. 17, and Fig. 18. with 10MB file size.

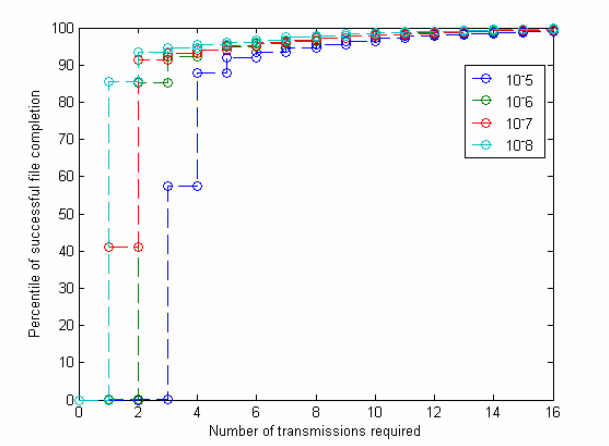


Figure 17. 10MB file latency performance with bad weather BER at 10^-3 with data obtained from minimum sampling.

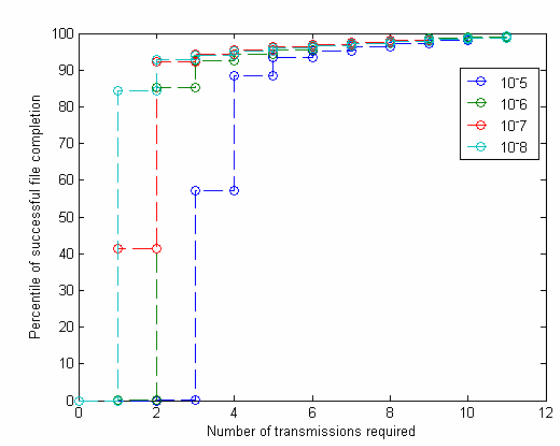


Figure 18. 10MB file latency performance with bad weather BER at 10^-4 with data obtained from minimum sampling.

The minimum sampling method requires 2 to 3 less transmissions than the result obtained from the average sampling method shown in Fig. 17, and, Fig. 18. More number of good weather data in the minimum sampling method results the improved delay performance than the results from other sampling methods.

The average number of transmissions required to complete the 99% file transmissions are recorded in Fig. 19. for different BER pairs. On average, at most 3.7 transmissions are required to transmit the 10MB file for the BER value pair, (10^-5, 10^-3), and 1.4 transmissions for the BER pair (10^-8, 10^-4). Differences between the average number of transmissions required from the average sampling method and the minimum sampling method are recorded in the Table 8.

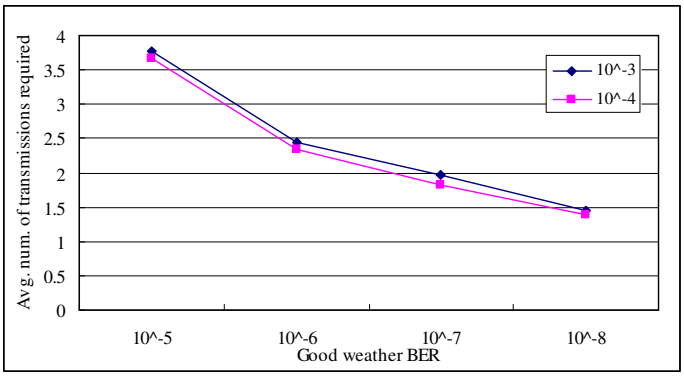


Figure 19. Average number of transmissions required to achieve 99% completion of the 10 MB file on the data collected from minimum sampling method.

Table 8. Differences between the average numbers of transmissions required to achieve the 99% completion of the 10 MB file with the average sampling and minimum sampling method.

BER	10^-5	10^-6	10^-7	10^-8
10^-3	0.4167	0.4755	0.2685	0.2771
10^-4	0.2612	0.1995	0.1922	0.1445

The overall CDF of three different sampling methods are compared in Fig. 20. through Fig. 23. for various BER pairs. It is found that the performance of the average sampling method is bounded between the maximum and the minimum sampling method. Due to the high weather correlation, the maximum transmissions requirements to complete the 99% of files are highly affected by the choice of sampling methods. However, they are the upper bounds and the actual average file transmissions requirements are significantly less.

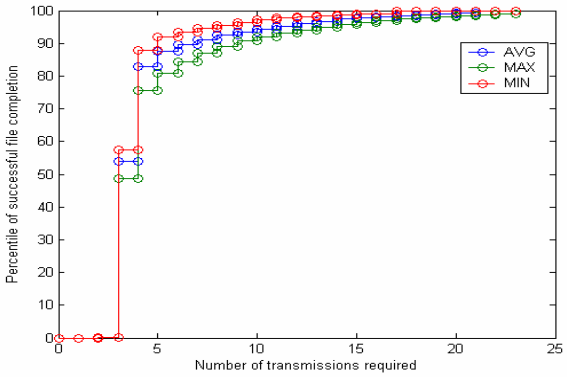


Figure 20. CDF of 10MB file latency performance with (10^-5,10^-3) BER on data obtained from the averaging, maximum, and minimum sampling method.

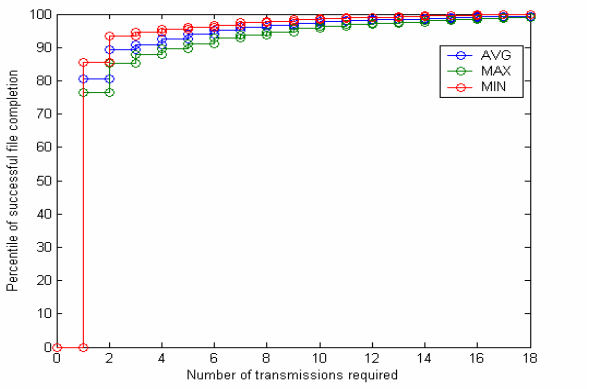


Figure 21. CDF of 10MB file latency performance with (10^-8,10^-3) BER on data obtained from the averaging, maximum, and minimum sampling method.

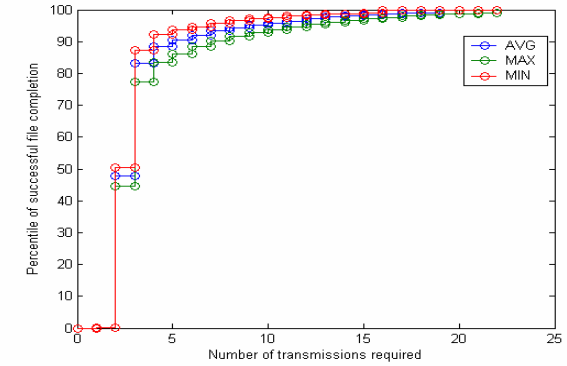


Figure 22. 1MB file latency performance with (10^-5,10^-3) BER with data obtained from averaging, maximum, and minimum sampling.

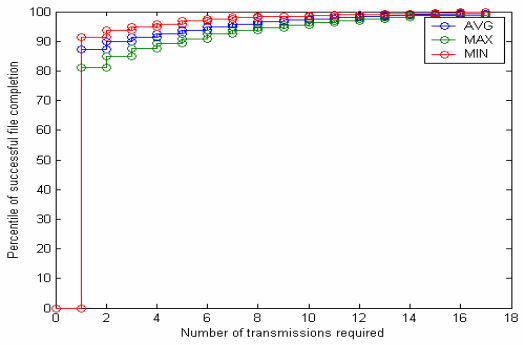


Figure 23. 1MB file latency performance with (10^-8,10^-3) BER with data obtained from averaging, choosing maximum, and minimum sampling.

D. Single-state weather channel (independent channel)

The goal of this section is to check the performance consistency with the previous work done by [2] and compare the performance of our two-state correlated weather channel with the single-state independent weather channel. In Appendix, the analytical expression of single-state channel model is derived from the two-state channel model by setting an equal BER value for good and bad weather state.

In this section, simulations are performed with 1.2x10^-5, 1.2x10^-6, 1.2x10^-7, and 1.2x10^-8 BER for the 10MB and 1MB file size. BER values are chosen to be consistent with [2]. We implicitly assume that 1.2x10^-5, 1.2x10^-6, 1.2x10^-7, and 1.2x10^-8 BERs are equivalent to 1.0x10^-5, 1.0x10^-6, 1.0x10^-7, and 1.0x10^-8 BERs for a comparison purpose. Figure 23. and Fig. 24. show the performance of the single-state channel model for 10MB and 1MB file size.

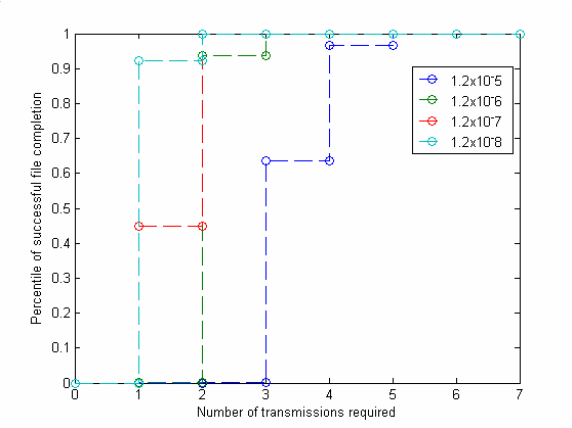


Figure 24. 10MB file latency performance with 1.2 x 10^-5, 1.2 x 10^-6, 1.2 x 10^-7, and 1.2 x 10^-8 BER in the single-state weather channel

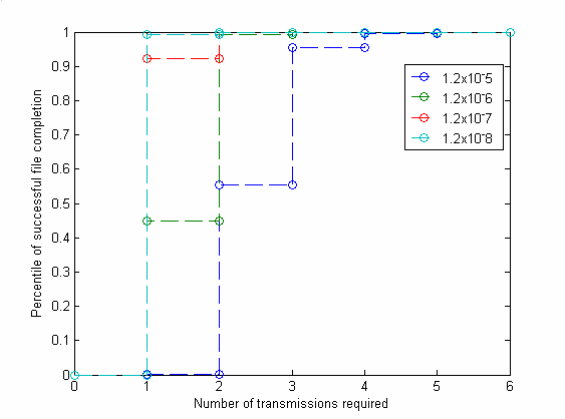


Figure 25. 1MB file latency performance with 1.2 x 10^-5, 1.2 x 10^-6, 1.2 x 10^-7, and 1.2 x 10^-8 BER in the single-state weather channel

The average numbers of spurts required are recorded in Table 9. From Table 9, we can see that the average delay in the correlated channel is slightly longer than the single-state weather model by at most 2 transmissions. In the single-state weather channel model with the best weather condition, approximately one transmission is enough to transmit 1MB file size.

Table 9. 10MB and 1MB file latency performance at 1.2 x 10^-5, 1.2 x 10^-6, 1.2 x 10^-7, and 1.2 x 10^-8 with the single state weather channel.

File Size	1.2 x 10^-5	1.2x 10^-6	1.2 x 10^-7	1.2 x 10^-8
10MB	3.3998	2.0617	1.5513	1.0769
1MB	2.4952	1.5771	1.077	1.008

Figure 26. and Fig. 27. show the overall performance comparisons between the BER of two-state channel model and the single-state weather channel model for 10MB and 1MB file sizes. We can see that the maximum numbers of transmissions required for the single weather state channel are far less than the two weather state channel model. Notice that the differences in the single-state and two-state models are the absence of correlation in the former. The result of incorporating weather correlation via the two-state model introduces the significant increase in the 99 percentile latency, which is mostly driven by tail events, while the average performance shows less dependency on correlation given the link availability over the long term is fixed.

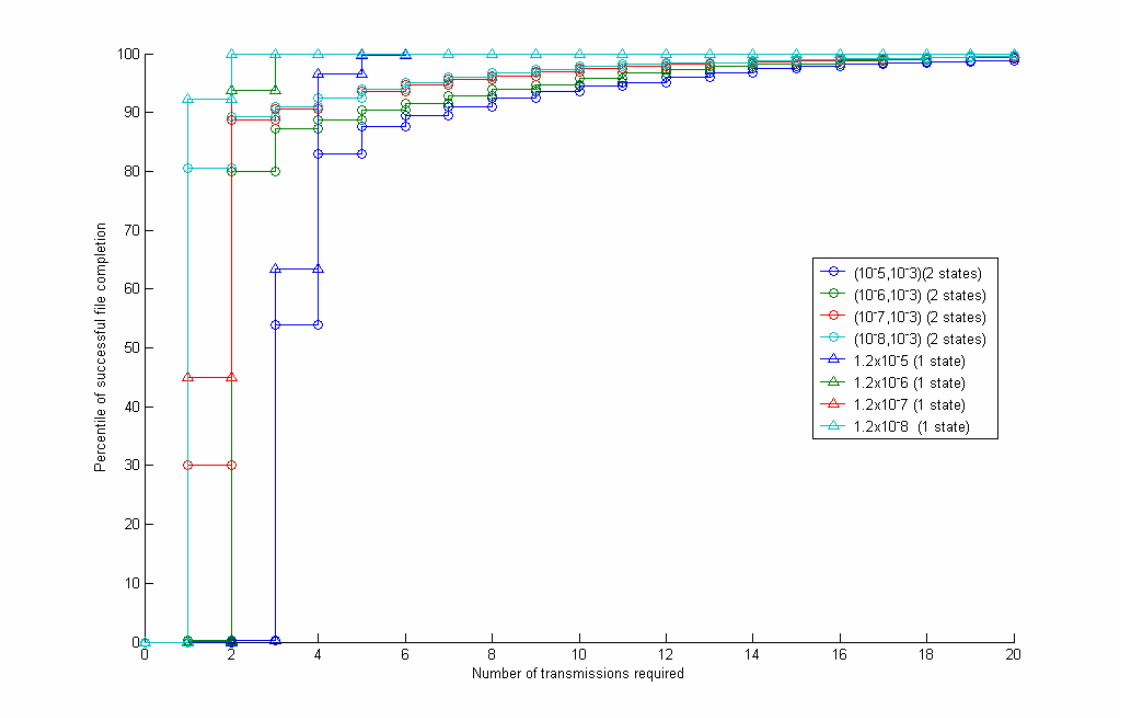


Figure 26. 10 MB file latency performance under the single-state weather channel and two-state weather channel model

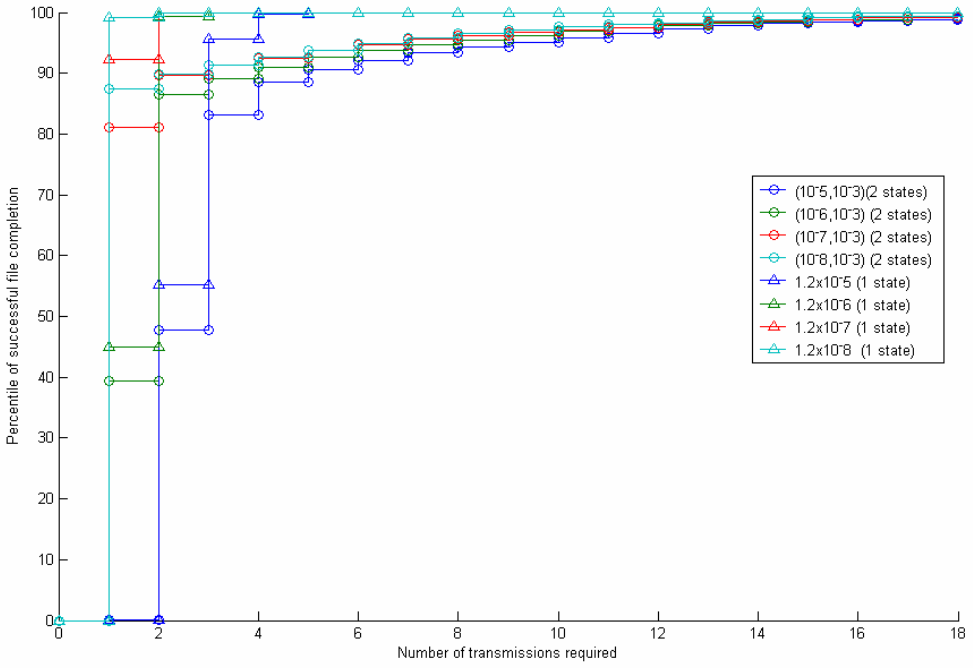


Figure 27. 1 MB file latency performance under the single-state weather channel and two-state weather channel

IX. Discussion

In this work, we captured the weather correlation via a two-state Markov chain. However, further work and improvements are needed using a high-order Markov chain to accurately model weather processes. Also, we sampled the data by averaging, selecting the maximum, and the minimum values. Alternatives ways of sampling should be explored in the future to improve our analysis. The CFDP evaluation based on the seasonal statistic, instead of whole data statistic, will provide more meaningful insight of the CFDP performance. Furthermore, evaluating the CFDP with higher weather availabilities will be the extension of the present study. It is anticipated that the higher weather availability will improve the average file transmission performance due to the shorter successive bad weather conditions. Therefore, in the higher weather availability, it is expected that the average file latency performance from the two-state Markov process in our work will more closely capture the true CFDP performance, because of less variations of bad weather patterns.

X. Conclusion

In this study, we present the framework to evaluate the CFDP performance under good and bad weather conditions. The CDF of the number of transmissions required to complete the file transfer for both 99 percentile latency and average latency are derived. The BERs at good and bad weather conditions affect the latency performance significantly, and varying file size has a moderate impact on the delay performance. While file size and BER affect the average latency, the 99 percentile latency is more dependent on the correlation of the weather patterns, which demonstrates the importance of developing weather mitigation strategy besides using retransmission protocols. Also, a fundamental improvement in the availability of the Ka-band link is expected to improve not just the average performance but significantly reduce the maximum number of transmissions required in the CDF plot.

Appendix

Verifying with the single weather channel

The validity of Eq. (15) can be checked with the single weather state case, where channel does not have a memory [2]. For this case, there is only one error state. We set, $p = q$, in Eq. (14) to represent the single weather state. Then, we have

$$f(k)g(k+1) = p. \tag{22}$$

If we substitute Eq. (21) into Eq. (15), following equation can be obtained:

$$\begin{aligned}
&= \left\{P(1,1) + P(2,1)... + P(i,1) + P(i+1,1) + ... + P(m,1)\right\}^N \\
&= \left\{P(1,1) + P(1,1)f(1)g(2) + P(1,1)f(1)g(2)f(2)g(3) + ...\right\}^N \\
&= \left\{(1-p) + p(1-p) + p^2(1-p) + ... + p^m(1-p)\right\}^N \\
&= \left\{\sum_{j=1}^{m} p^{j-1}(1-p)\right\}^N \\
&= (1-p^m)^N.
\end{aligned} \tag{23}$$

Therefore,

$$\begin{aligned}
P[N_s \le m] &= \sum_{w} P[N_s \le m \mid w_k] \cdot P[w_k] \\
&= (1-p^m)^N.
\end{aligned} \tag{24}$$

For single weather state channel model, there is only one type of weather sequence. Hence, *n*, and, $P[w_k] = 1$. This Eq. (23) is the same as the analytical expression derived from the independent random error channel model in [2]. This completes the proof.

Acknowledgments

The authors would like to thank Dr. Owhadi at California Institute of Technology for introducing Hoeffding's Inequality and other invaluable comments on Section VII. Also, the authors would like to acknowledge fellow researchers in the Communications Research Section at the Jet Propulsion Laboratory for their suggestions and feedback, which greatly clarified and improved this work. The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1]CCSDS File Delivery Protocol, CCSDS 727.0-B-2, URL: http://ccsds.org/

[2]Gao, J.L., and Segui, J.S., "Performance Evaluation of the CCSDS File Delivery Protocol – Latency and Storage Requirement", *IEEE Aerospace Conference.* 2004.

[3] W. Li, C. L. Law, V.K. Dubey, and J.T. Ong, "Ka-band land mobile satellite channel model incorporating weather effects", *IEEE Communications Letters*, 5. 2001.

[4] V.Y.Y.Chu, P.Sweeney, J.Paffett, M.N.Sweeting, "Characterizing error sequences of the low earth orbit satellite channel and optimization with hybrid-ARQ schemes", Proc. *IEEE GLOBECOM*, Sydney, pp.3405-3410, 1998.

[5] Sun, J., Gao J., Shambayatti, S., and Modiano, E., "Ka-Band Link Optimization with Rate Adaptation," *IEEE Aerospace Conference*, March, 2006.

[6] Sklar, "Digital Communications: Fundamentals and Applications*",* 2nd ed., Prentice Hall, 2nd, 2001.

[7] Gilbert, E. N., "Capacity of a Burst-Noise Channel", *Bell Sys. Tech. J.*, 39,pp. 1253-1266, Monograph 3683, September 1960.

[8] Elbert, B.," Introduction to Satellite Communications", 2nd ed., Artech House, 1998.

[9] Garcia, L.," Introduction to probability for electrical and computer engineers", 2nd ed., Addison Wesley, 2nd, 1993.

[10] Hoeffding W.**, "**Probability inequalities for sums of bounded random variables", *American Statistical Association Journal*, pages 13-30, 1960.

[11] Feyerherm A., and Bark, L., "Goodness of Fit of a Markov Chain Model for Sequences of Wet and Dry Days**"**, *Journal of Applied Meteorology*, Vol. 6, No. 5, pp. 770–773.1967.

[12] Jiao, C., Schwiebert, L., and Xu, B., "On Modeling the Packet Error Statistic in Bursty Channels**"**, *IEEE Local Computer Networks Conference*, November, 2002.